

## DOCUMENT RESUME

ED 317 207

IR 053 020

AUTHOR Atkinson, Steven D.; Walker, Geraldene  
TITLE Online Access in the Humanities: Implications for  
Researchers. A Report to the Council on Library  
Resources.  
INSTITUTION Council on Library Resources, Inc., Washington,  
D.C.  
PUB DATE 25 Sep 89  
NOTE 34p.  
PUB TYPE Reports - Research/Technical (143)  
  
EDRS PRICE MF01/PC02 Plus Postage.  
DESCRIPTORS Art; \*Bibliographic Databases; Comparative Analysis;  
Content Analysis; History; \*Humanities; \*Information  
Sources; Literature; Music; \*Online Searching;  
\*Research Tools; \*Search Strategies  
IDENTIFIERS Boolean Operators; Natural Language

## ABSTRACT

Despite the growth in number of online bibliographic databases available to assist scholars seeking information in the humanities, it remains a matter of concern to librarians and information professionals that these research tools are not as widely used as they might be. This report surveys a selected group of online databases (i.e., "America: History and Life," "Arts and Humanities Search," "Art Literature International," "Artbibliographies Modern," "Historical Abstracts," "Linguistics and Language Behavior Abstracts," "MLA Bibliography," "Philosopher's Index," and "Religion Index") to identify conceptual relationships between the different disciplines (i.e., art, history, literature, music, and interdisciplinary studies) in the humanities. A comparison is made between the effectiveness of natural language and controlled vocabulary for maximizing recall and the degree of uniqueness of records retrieved from the various files using four search types: (1) single subject terms of a specific nature; (2) single subject terms of a generic nature; (3) subject phrases; and (4) single subject terms combined with the Boolean "AND." The results demonstrate the possibilities for more productive use of online bibliographic databases as a resource for scholarly research in the humanities. Nine tables present the results analyses of the data. (34 references) (Author/SD)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED317207

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

\* This document has been reproduced as  
received from the person or organization  
originating it.

□ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

## Online Access in the Humanities: Implications for Researchers

### A Report to the Council on Library Resources

Submitted

by

Steven D. Atkinson  
Assistant Coordinator, Computer Search Service  
University Libraries  
The University at Albany  
State University of New York at Albany  
1400 Washington Ave.  
Albany NY 12222

and

Dr. Geraldene Walker  
Assistant Professor  
School of Information Science and Policy  
University at Albany  
State University of New York at Albany  
1400 Washington Ave.  
Albany NY 12222

September 25, 1989

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

Steven D. Atkinson

IR053020

## INTRODUCTION

Despite the growth in the number of online files available to assist scholars seeking information in the humanities, it remains a matter of concern to librarians and information professionals that these resources are not as widely used as they might be. More widespread acceptance of the computer as a tool for scholarly research has led to the development of an extraordinarily wide variety of online databases containing references to primary and secondary research materials, but their use has been limited. International links among humanities scholars are still rudimentary (1), and most of them use the computer mainly for word processing, text analysis and desktop publishing, rather than regarding it as a tool for information-gathering. The current low levels of online use by humanists are a result not only of their traditional information-seeking styles, but also of the nature of the subject fields and the coverage provided by the online files.

The investigation reported here uses a selected group of bibliographic files to identify conceptual relationships between the different subject disciplines within the field. It also compares the effectiveness of natural language and controlled vocabulary for maximizing recall and the degree of uniqueness of records retrieved from different files. An overview of this type, demonstrates the possibilities for the more productive use of online bibliographic databases as a resource for scholarly research in the humanities.

## BACKGROUND

Research over the past twenty or thirty years has provided a picture of the information needs and information-seeking behavior of both scientists (2, 3) and social scientists (4, 5) and of the differences between users in different disciplines (6, 7). Recent studies have also investigated the use of online sources and have emphasized the importance of multifile searching in a range of scientific fields in order to provide adequate coverage of the literature (8, 9, 10, 11, 12).

While humanities scholars may be expected to differ from those in other fields in their information needs, information-seeking behavior and information use, existing studies of these differences have been largely descriptive and generally restricted to a single discipline. Stebelman considers the "admixture of indifference, skepticism, and in some cases, borderline hostility" of humanistic scholars towards online sources to be due to "psychological blocks and philosophical reservations." (13) The objective knowledge of the humanistic disciplines certainly does appear to have characteristics different from those of the sciences (14), and their concepts and vocabulary do not have the same logical clarity as those of scientists. Humanistic knowledge is more open-ended, requiring complex philosophical and aesthetic judgments, and their disciplines are not normally organized in the hierarchical fashion of the sciences.

. Current online files have a number of limitations from the point of view of the humanist, particularly in terms of coverage. The

humanities databases were late arrivals on the online scene and thus cover the periodical literature for a limited number of years. This restricted coverage is a considerable drawback, since humanistic scholarship has strong historical dimensions, such that books are at least as important as journal material, and retrospective coverage even more vital than currency. Despite the fact that many topics in the humanities are obviously interdisciplinary, it has been pointed out that most university departments are notoriously insular (15) and that few scholars are aware of the major indexes and abstract services outside their own disciplines (16). Indeed, it has been suggested that historians, for instance, may find almost all indexes and abstracts 'irrelevant' (17), and in general their use of online services has been inhibited by a typical "resistance to new modes of information access" (18).

Most of what has been published on the information use of humanities scholars is subjective (19), and those few research studies which do exist provide analyses of only single disciplines in isolation (20, 21, 22). This investigation emphasizes the connections between different subject fields, so as to show the importance of interdisciplinary links, which can now be more easily utilized through the use of online bibliographic sources. The information provided by this study will become increasingly important as the implementation of scholars' workstations facilitate the growth of interdisciplinary research.

## METHODOLOGY

The subject 'profile' approach pioneered by Williams (23) and recommended by Tenopir (24) was adopted to evaluate the database coverage of a series of topics across a range of subject fields within the humanities. The aim was to identify the scatter of topics and the overlap of subject terminology and records between files. Although this method is partly dependent on database indexing policies, it provides useful information regarding the makeup of the 'core' of a subject field and eliminates the subjectivity that inhibits most other evaluative approaches.

Subject terms for a variety of search topics were searched across nine humanities files available on the DIALOG system:

America History and Life,  
Arts and Humanities Search,  
Art Literature International,  
Artbibliographies Modern,  
Historical Abstracts,  
Linguistics and Language Behavior Abstracts,  
MLA Bibliography,  
Philosopher's Index, and  
Religion Index.

The methodology used involved the selection of a list of terms that were designed to represent different subject 'types' and the execution of the searches across all the databases being investigated. Crucial to its effectiveness is the

initial choice of topics for searching. It was hypothesized that the level of specificity would vary by subject field, even within areas like the humanities, where the evolution of the vocabulary is slow and a variety of quasi-synonymous terms may be used to express the same concept. Topics covering a range of subject fields were thus selected to represent different levels of specificity based initially on Wiberly's categories of humanistic vocabulary (25). He identifies four groups:

1. singular proper terms -- the names of unique persons or single creative works;
2. enumerable proper terms -- a collective group which may be completely enumerated;
3. general proper terms -- often difficult to define and covering a range of meanings and types, and
4. common terms -- any one of a class of things or the class itself.

Although these categories do vary in level of specificity, they include many proper nouns, which are relatively straightforward to search. It was decided that this research would concentrate on subject search terms, rather than proper names, so Wiberly's categories were adapted to provide the following classification of four search 'types', each at two levels of specificity:

1. single subject terms of a specific nature (discipline-specific terms such as WATERCOLOR or JAZZ, and interdisciplinary terms such as CENSORSHIP);

2. single subject terms of a generic nature (discipline-specific terms such as SURREALISM, or IMPERIALISM and interdisciplinary terms such as PSYCHOANALYSIS);
3. subject phrases (discipline-specific phrases such as DIVINE RIGHT, or RESTORATION COMEDY, and interdisciplinary phrases such as POPULAR CULTURE);
4. single subject terms combined with Boolean AND (discipline-specific combinations such as SERFS AND RUSSIA, or COMPUTERS AND COMPOSITION, and interdisciplinary combinations such as IRONY AND HUMOR)

Search terms were also selected to try to represent the diversity of the field. First, from four of the major disciplines of the humanities -- Art, History, Literature and Music -- and then two groups of topics considered to be 'interdisciplinary'. All search terms were tested for typicality by scholars in the appropriate disciplines. Three of these disciplinary areas are each directly represented by DIALOG databases -- Art by Artbibliographies Modern and Art Literature International (RILA), History by America: History and Life and Historical Abstracts, and Literature by MLA Bibliography and more marginally by Linguistics and Language Behavior Abstracts (LLBA). Two of the other files used -- Religion Index and Philosopher's Index -- were not represented by discipline-specific search terms and one major discipline -- Music -- was included, though not searched in its major file (Repertoire de Musique). Arts & Humanities Search was also included to represent interdisciplinary coverage of the humanities, and its performance was of particular interest for comparison with the discipline-specific files.

Most terms were truncated to allow for variant spellings and word endings and were searched only on the title (TI) and descriptor (DE) fields to maximize specificity, though no attempt was made to search synonymous terms. Although the use of truncation may lead to some false drops (e.g. WITCHITA for WITCH?), there is no reason to suppose that it will effect one database more than another. Searches were limited to documents in English and to publication dates between the years 1983 and 1987 in order to restrict output to manageable size and to standardize file coverage. The result was a four by six matrix of subject topics which were searched across all nine databases -- 216 search profiles in all (see Table 1).

Table 1: Terms searched by type and subject field

|                   | SINGLE TERM<br>SPECIFIC | SINGLE TERM<br>GENERIC | PHRASES                 | COMBINED<br>TERMS               |
|-------------------|-------------------------|------------------------|-------------------------|---------------------------------|
| ART               | watercolo?              | surreal?               | art()deco               | cat AND<br>symbol?              |
| HISTORY           | witch?                  | imperial?              | divine()right           | serf AND<br>russia?             |
| LITERAT.          | picaresque              | romantic?              | restoration()<br>comedy | computer?<br>AND<br>composition |
| MUSIC             | jazz                    | improvisat?            | gregorian()<br>chant    | creativ?<br>AND<br>imagin?      |
| INTER-<br>DISCIP. | sexual?                 | marriage?              | feminist()<br>crit?     | magic?<br>AND<br>folklore       |
| INTER-<br>DISCIP. | censorship              | psychoanal?            | popular()<br>culture    | irony<br>AND<br>humo??          |

## EVALUATION

Evaluation is based on the assumption that so long as the standard of searching is consistent, then postings figures can be regarded as an indication of search effectiveness. Although early research had suggested that an inverse relationship existed between precision and recall (26), a more recent study found that higher recall was positively related to larger numbers of relevant records and vice versa (27). It is therefore presumed for this investigation that retrieved sets which are larger may be expected to contain a greater number of relevant citations. It is recognized that on any search it is possible to improve recall at the expense of relevance, but the search strategies used made no attempt to maximize postings by including alternative synonymous terms. The same strategy was used for all search queries with only the search terms changed.

A detailed analysis of the output for all searches enabled the identification for each search topic of:

1. the contribution of each database to total postings for the different subjects (scatter by subject field);
2. the contribution of unique records by database and by term type (duplication);
3. the share contributed to retrieval by natural language and controlled vocabulary and the overlap between the two (terminology); and
4. the contribution of each database to total postings for the different term 'types' (scatter by search type).

## SCATTER BY SUBJECT FIELD

'Scatter' is the term used to identify the spread of terms, both subject terms and term 'types', across the databases. For example, what percentage of the material retrieved on art-related subjects was provided by Artbibliographies Modern and RILA, the obvious search files? Previous experience had suggested that a considerable amount of art history material is included in Historical Abstracts (28), for example, but what about the other humanities files? Information of this type is particularly helpful for novice searchers, who may have only limited experience of databases in their own fields and none at all of those in other fields. The results of this scatter analysis are presented in Table 2.

Table 2: Database Scatter by Subject Area of Search Terms

| Databases  | Search Subjects |         |       |       |        |        |
|------------|-----------------|---------|-------|-------|--------|--------|
|            | ART             | HISTORY | LIT.  | MUSIC | INTER1 | INTER2 |
| Amer.Hist. | 2.3%            | 24.4%   | 7.5%  | 6.7%  | 27.5%  | 31.6%  |
| Hist.Abs.  | 2.8%            | 33.7%   | 11.0% | 0.9%  | 19.4%  | 32.1%  |
| Art. Mod.  | 78.6%           | 1.1%    | 11.7% | 1.4%  | 3.2%   | 4.1%   |
| RILA       | 42.8%           | 7.6%    | 15.1% | 0.7%  | 12.9%  | 20.9%  |
| LLBA       | 0.7%            | 1.4%    | 0.4%  | 7.7%  | 75.7%  | 14.1%  |
| MLA        | 3.1%            | 4.4%    | 26.6% | 2.3%  | 35.7%  | 27.9%  |
| Phil.Ind.  | 2.6%            | 4.8%    | 12.7% | 1.6%  | 43.4%  | 34.9%  |
| Rel.Ind.   | 0.1%            | 9.1%    | 6.2%  | 0.5%  | 67.8%  | 16.3%  |
| Art & Hum. | 5.0%            | 22.0%   | 20.6% | 11.6% | 24.7%  | 16.1%  |

Initial inspection of this table suggests that the expected concentrations of postings do seem to occur (e.g. Historical Abstracts has the highest percentage of postings for search terms in the area of history, Artbibliographies Modern has the largest for Art, etc.), but all files also provide at least some postings for every search topic. America: History & Life is an exception to this pattern and would appear to provide a more general coverage than history alone. MLA and Arts & Humanities Search appear to be major sources for material in all subject fields, particularly for history and literature searches. In fact, a comparison of postings from A&HS with those from all other databases highlights the importance of this file (even though it is limited by its lack of assigned indexing terms), with its contributions ranging from 33% to 82% of overall citations retrieved for a single topic.

Since the lack of detail made it difficult to identify any over-riding pattern in this distribution, a subset was developed for more detailed analysis by grouping the related pairs of disciplinary files (Artbibliographies Modern and RILA, Historical Abstracts and America:History and Life, and LLBA and MLA) and identifying their performance on their subject-related search terms. This concatenation produced the matrix presented in Table 3, which shows more clearly how subject terms are indeed most highly posted in their linked subject databases. History is again an exception to this pattern, probably due to the generality influence of the America: History & Life file mentioned previously.

Table 3: Distribution of Subject Terms by Database Groupings

| Database Groupings              | Search Subjects |         |          |           | OVERALL |
|---------------------------------|-----------------|---------|----------|-----------|---------|
|                                 | ART             | HISTORY | LITERAT. | INTERDIS. |         |
| ART<br>(Art. Mod.<br>& RILA)    | 50.9%           | 1.3%    | 3.9%     | 3.0%      | 8.5%    |
| HISTORY<br>(AHL &<br>Hist. Ab.) | 4.4%            | 23.3%   | 6.2%     | 18.7%     | 14.3%   |
| LITERATURE<br>(MLA & LLBA)      | 11.5%           | 7.7%    | 37.6%    | 39.1%     | 27.5%   |
| INTERDIS.<br>(A&HS)             | 33.2%           | 67.7%   | 52.3%    | 39.1%     | 49.7%   |

A chi-square analysis ( $df = 9$ ,  $\alpha = 0.05$ ) confirms that a significant relationship exists between the subjects of the queries and the subject areas covered by the databases. Thus it can be states that although there is obviously a lot of material in other files, the major discipline-based files of the humanities are the major sources for subject searches in their own fields.

The interdisciplinary nature of MLA Bibliography and the contribution of A&HS, particularly to history and literature searches, are worth noting. They are obviously both important sources for any subject field within the humanities. It seems clear that any search requiring comprehensive coverage of a single subject field in the humanities needs to searched across a whole range of files.

#### DUPLICATION

The next question to be addressed is the level of overlap between the online files in terms of retrieved records. In other words, how much of this scattered material is new and how much merely duplication? Documentation for the various databases suggest that only minimal overlap between files in terms of individual records should be expected, and that searching additional files is likely to contribute mainly new citations. An indication of the percentage of unique records contributed by each database for a selection of the search terms is presented in Table 4.

Table 4: Sample of unique records by database

|           | jazz  | picaresque | art deco | divine right |
|-----------|-------|------------|----------|--------------|
| Amer.Hist | 16%   | 0          | 3.2%     | 0            |
| Hist.Ab.  | 4.9%  | 1.3%       | 4.8%     | 0            |
| Art. Mod. | 2.8%  | 0          | 30.2%    | 0            |
| RILA      | 0.7%  | 0          | 7.9%     | 0            |
| LLBA      | 0.7%  | 0          | 0        | 0            |
| MLA       | 29.2% | 38.5%      | 0        | 12.5%        |
| Phil.Ind. | 0     | 0          | 0        | 0            |
| Rel.Ind.  | 2.8%  | 0          | 0        | 31.3%        |
| A&HS      | 39.6% | 44.9%      | 44.5%    | 37.5%        |

These few examples show that quite often often unexpected databases provide not only postings, but unique citations and emphasize the need for multifile searching. It is interesting to note that, despite the fact that it has no added subject terms from a controlled vocabulary, A&HS provides additional citations for nearly any search topic in the field of the humanities.

The duplication of records between files would appear to be minimal overall, with only 8.7% of records appearing in more than one file. In order to determine if differences existed between subject groups and individual databases, a more detailed analysis produced the results displayed in Tables 5 and 6.

Table 5: Percentage overlap of records within subject groups

| Search terms      | Unique records | Duplicates |
|-------------------|----------------|------------|
| HISTORY           | 94.0%          | 6.0%       |
| ART               | 72.8%          | 27.2%      |
| LITERATURE        | 91.5%          | 8.5%       |
| MUSIC             | 95.3%          | 4.7%       |
| INTERDISCIPLINARY | 94.9%          | 5.1%       |
| AVERAGE           | 91.3%          | 8.7%       |

These figures confirm that in general overlap between files is remarkably low. It appears that art is the most dispersed of the subject fields investigated here, with much higher levels of duplication than the other fields. This duplication is largely the result of overlap between the two major art databases (Artbibliographies Modern and RILA). A matched pairs t-test ( $df = 4$ ,  $\alpha = 0.05$ ) confirmed that the difference between art and the other fields in terms of overlap is significant.

Table 6: Contribution of unique records by database

|            | Unique records |
|------------|----------------|
| Amer.Hist, | 7.05%          |
| Hist.Abs.  | 11.5%          |
| Art. Mod.  | 4.8%           |
| RILA       | 3.8%           |
| LLBA       | 2.9%           |
| MLA        | 19.9%          |
| Phil.Ind.  | 5.2%           |
| Rel.Ind.   | 11.7%          |
| A&HS       | 27.3%          |

In all the search strategies LLBA and the two art files (Artbibliographies Modern and RILA) have the lowest numbers of unique records, suggesting that their coverage duplicates much material found in other databases. The most interdisciplinary of the files -- MLA and A&HS -- have the highest proportion of unique records not available elsewhere, with Historical Abstracts also having a large percentage of unique records. A future investigation is needed to see how these figures relate to the journal coverage of the different databases.

#### SEARCH TERMINOLOGY

The third area of interest was the comparative effectiveness of natural language and controlled vocabulary for information retrieval for different subject fields and types of search. This question involved the separation of postings figures for the title (TI) and descriptor (DE) fields, representing natural language and controlled vocabulary, respectively. It is generally accepted among professional searchers that both types of terminology are necessary for maximum retrieval, and it has been suggested that controlled vocabulary is especially effective for improving recall (29). Although the indexing vocabularies of the different files vary, they can be used as the means for a broad assessment of the improvement in retrieval to be attained by adding thesaural terms to the natural language of a user's search query.

A number of authors have discussed the comparative advantages of free-text and controlled vocabulary for online retrieval (30, 31, 32). They have pointed out that although free-text assists

direct user access by providing simplified searching, it also places the burden for success entirely on the imagination and ingenuity of the searcher. It has been demonstrated that the selection effective good search terms is the greatest source of failure for novice users (33). This finding suggests that the newly-developing frontends and expert systems will need to include facilities to assist the 'enhancement' of a user's natural language search vocabulary by providing a choice of related and synonymous terms from an online thesaurus.

The question addressed here concerned the differences in postings to be achieved when using controlled vocabulary as compared with natural language and the levels of overlap between the two. How many postings, in other words, were unique to one type of search key? These findings are presented in two ways -- by search term (see Table 7), and by database (see Table 8). (A&HS is excluded from this analysis because it has no controlled vocabulary.

Table 7: Comparison of natural language and controlled vocabulary by search term

|                      | Nat. Lang. | Controlled | Both  |
|----------------------|------------|------------|-------|
| SPECIFIC TERMS       |            |            |       |
| watercolor?          | 4.6%       | 75.9%      | 19.7% |
| witch?               | 10.8%      | 39.8%      | 49.4% |
| picaresque           | 10.0%      | 40.0%      | 40.0% |
| jazz                 | 23.6%      | 36.6%      | 39.8% |
| mean                 | 12.3%      | 48.1%      | 37.2% |
| GENERIC TERMS        |            |            |       |
| surreal?             | 7.4%       | 52.4%      | 40.2% |
| imperial?            | 40.7%      | 35.4%      | 23.9% |
| romantic?            | 23.5%      | 34.9%      | 41.6% |
| improvisat?          | 54.5%      | 18.2%      | 27.3% |
| average              | 31.5%      | 35.2%      | 33.3% |
| PHRASES              |            |            |       |
| art()deco            | 0          | 50.0%      | 50.0% |
| divine()right        | 21.4%      | 57.1%      | 21.5% |
| restoration()comed?  | 71.4%      | 21.4%      | 7.2%  |
| gregorian()chant?    | 50.0%      | 25.0%      | 25.0% |
| average              | 35.7%      | 38.4%      | 25.9% |
| COMBINED TERMS       |            |            |       |
| cat? AND symbol?     | 0          | 100.0%     | 0     |
| serf? AND russia?    | 0          | 10.0%      | 90.0% |
| comput? AND compos?  | 8.3%       | 83.3%      | 8.3%  |
| creativ? AND imagin? | 40.7%      | 48.1%      | 11.1% |
| average              | 12.3%      | 60.4%      | 27.4% |
| Overall Average      | 22.9%      | 45.5%      | 30.9% |

These figures show that, on average, the controlled vocabulary retrieved significantly higher postings (45.5%) than natural language (22.9%). Although levels of overlap vary (from zero to 90%), they are usually high (30.9%) and do not appear to be affected by type of search term. Natural language performed almost as well as controlled vocabulary for generic terms and phrases, while the descriptor field was most successful field for specific terms and combined terms. Since different files use different controlled vocabularies, it was thought possible that these patterns might vary by database. Table 8 shows the differences between natural language and controlled vocabulary divided by database.

Table 8: Comparison of natural language and controlled vocabulary by database (postings)

| DATABASE   | Nat.Lang. | Controlled | Both  |
|------------|-----------|------------|-------|
| Amer.Hist. | 21.1%     | 48.2%      | 30.7% |
| Hist.Abs.  | 21.4%     | 44.1%      | 34.5% |
| Art. Mod.  | 12%       | 64.3%      | 23.7% |
| RILA       | 4%        | 68%        | 27.9% |
| MLA        | 13.7%     | 56.9%      | 29.4% |
| LLBA       | 6%        | 90.8%      | 3.2%  |
| Phil.Ind.  | 18.5%     | 45.1%      | 36.4% |
| Rel.Ind.   | 16.9%     | 56.7%      | 26.4% |
| Average    | 15.2%     | 56.4%      | 28.4% |

Although the controlled vocabulary achieved higher postings on all databases, such apparent differences are not necessarily significant in statistical terms. A matched pairs t-test ( $df = 7$ ,  $\alpha = 0.05$ ), however, produced a t-value of -2.29, which confirmed that these differences are in fact significant. Controlled vocabulary performs best in the fields of art and literature, with a particularly noteworthy performance on LLBA. These findings suggest noteworthy differences, though whether they are due to the nature of the controlled vocabularies themselves, or a reflection of some subtle interdisciplinary linkages, could not be determined without further investigation.

Although the controlled vocabulary did retrieve higher postings than natural language in all the files investigated, such differences may be relatively unimportant in an intermediary searching environment in which searchers are aware of the value of both types of vocabulary. But they do have important implications for the training of end-user searchers, who are more likely to rely on natural language search terms and to be unaware of the limitations of such a search strategy.

#### SCATTER BY SEARCH TYPE

The next area of interest -- scatter by type of search term -- is related to both database coverage and search terminology.

Inspection of Table 9 shows no obvious relationships, though it appears that discipline-specific files do not necessarily perform best for any one particular type of term. In general, single terms (either specific or generic) retrieve

higher postings than phrases or combined (ANDed) terms. This result is not surprising, and no statistical relationships were identified.

Table 9: Database Scatter by Term Classification

| Database   | Type of Term               |                           |         |                   | OVERALL |
|------------|----------------------------|---------------------------|---------|-------------------|---------|
|            | SINGLE<br>TERM<br>SPECIFIC | SINGLE<br>TERM<br>GENERIC | PHRASES | COMBINED<br>TERMS |         |
| Amer.Hist. | 2.4%                       | 2.3%                      | 8.3%    | 0.5%              | 2.9%    |
| Hist.Abs.  | 4.6%                       | 8.4%                      | 18.0%   | 7.6%              | 8.0%    |
| Art. Mod.  | 5.9%                       | 2.0%                      | 2.2%    | 0.5%              | 3.3%    |
| RILA       | 2.4%                       | 1.5%                      | 4.0%    | 1.1%              | 2.0%    |
| LLBA       | 4.9%                       | 0.6%                      | 0       | 10.8%             | 2.1%    |
| MLA        | 23.7%                      | 26.0%                     | 24.4%   | 19.5%             | 25.0%   |
| Phil.Ind.  | 1.7%                       | 1.4%                      | 0.3%    | 0.5%              | 1.4%    |
| Rel.Ind.   | 9.1%                       | 11.1%                     | 8.3%    | 16.2%             | 10.3%   |
| A&HS       | 45.3%                      | 46.5%                     | 34.4%   | 43.2%             | 45.0%   |

The lower postings for combined (ANDed) terms, as compared with phrases, is somewhat surprising, particularly in view of the fact that ANDed combinations are frequently enhanced by wrong coordinations which occasion 'false drops' (the retrieval of irrelevant material). It appears that the phrasal terms may, in fact, be the best indicators of interdisciplinarity, since they are probably the most specific of the search keys used.

Once again the discipline-related files were grouped to test whether relationships existed between the database groups and the search 'types'. The computed chi-square ( $df = 9$   $\alpha = 0.05$ ) for these groups suggests that a significant relationship of some kind does exist, though its exact nature will require further investigation.

#### CONCLUSIONS

The methodology used for this research provides a relatively straightforward and inexpensive method for measuring and comparing the coverage of databases for different subject areas. It must be remembered, however, that this methodology is limited through being based solely on postings figures and taking no account of the relevance of the citations retrieved. It is based on the assumption that higher recall will also produce more relevant material. In addition, the implicit assumption that each citation and each search term are of equal importance is obviously an over-simplification, and the choice of search terms used here may not necessarily be representative. Despite these drawbacks, it is believed that online databases can provide a useful source of information regarding the spread of coverage for a given subject across different disciplines.

The results presented here give an indication of the effectiveness of each of the databases for retrieving information in different subject fields, at different levels of specificity and from using different types of search vocabulary. Access to information of this kind can assist searchers with the choice of appropriate databases and search terms for a given search topic. Subject relationships between the different disciplines within the field of the humanities appear to be more diverse and interconnected than the behavior of academic researchers had previously suggested, and the importance to multiple files is clear. Unfortunately, research on end-user access to online information has indicated that novice searchers tend to perform best using one system and a single database (34). The research reported here suggests that this approach, though undoubtedly simpler for user and trainer alike, will not lead to the most effective search results. Multi-file searching is a complicated process, and the challenge for information professionals is to simplify access for (possibly unenthusiastic) naive users by the development of training programs and software 'filters'. The challenge to system designers lies in the determination of the most effective division between explicit and transparent system features, so as best to represent the conceptual framework of the average untrained user. The potential of electronic research techniques for interdisciplinary and cross-disciplinary research and for the provision of a broad new synthesis of perspectives lies in making the available systems simple, convenient and easy to use.

## REFERENCES

1. Raben, Joseph. Electronic Scholars Resource Guide: The Humanities go online. Proceedings of the Eighth National Online Meeting, New York, May 3-7, 1989:409-411.
2. Allen, Thomas J. Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R & D Organization. Cambridge, MA: MIT Press, 1977.
3. Crane, Diana. Invisible Colleges: Diffusion of Knowledge in Scientific Communities. Chicago: University of Chicago Press, 1972.
4. Goldberg, A.L. "Information Needs of Social Scientists and Ways of Meeting Them." International Social Science Journal 23 (2) 1971: 273-284.
5. Line, Maurice B. "The Information Uses and Needs of Social Scientists: An Overview of Infross." ASLIB Proceedings 23 (8) 1971: 412-434.
6. Martyn, John and Slater, M. "Tests on Abstracts Journalism." Journal of Documentation 20 (4) December 1964: 215-235.
7. Martyn, John. "Tests on Abstracts Journals: Coverage, Overlap and Indexing." Journal of Documentation 23 (1) March 1967: 45-70.
8. Tenopir, Carol. "Distribution of Citations in Databases in a Multi-disciplinary Field." Online Review 6 (5) October 1982: 399-419.
9. Sharma, V.S. "Information Resources for Recreation, Leisure, and Sport." Database 5 (3) August 1982: 297-313.
10. Epstein, Barbara A. and Angier, Jennifer J. "Multi-database Searching in the Behavioral Sciences, Part 1: Basic Techniques and Core Databases." Database 3 (3) September 1980: 9-15.
11. Angier, Jennifer J. and Epstein, Barbara A. "Multi-database Searching in the Behavioral Sciences, Part 2: Special Applications." Database 3 (4) December 1980: 34-40.
12. Meyer, Daniel E. et al. "Comparison Study of Overlap Among 21 Scientific Databases in Searching Pesticide Information." Online Review 7 (1) January 1983: 33-43.
13. Stebelman, Scott D. "On-Line Searching and the Humanities: Relevance, Resistance, and Marketing Strategies." In Proceedings of the Second National Online Meeting. Compiled by Martha E. Williams and Thomas H. Hogan. Medford, NJ: Learned Information, Inc., 1981: 443-453

14. Garfield, Eugene. "Is Information Retrieval in the Arts and Humanities Inherently Different from that in Science? The Effect that ISI's Citation Index for the Arts and Humanities is Expected to have on Future Scholarship." Library Quarterly 50 (1) January 1980: 40-57.
15. Gombrich, E.H. "Research in the Humanities: Ideals and Idols." Daedalus 102 Spring 1973: 1-10.
16. Stebelman op. cit.
17. Stieg, Margaret F. "The Information Needs of Historians." College & Research Libraries 42 (6) November 1981: 549-560
18. Mackeasy, Eileen M. "A Perspective on Secondary Access Services in the Humanities." Journal of the American Society for Information Science 33 (3) May 1982: 146-151.
19. Weintraub, Karl J. "Humanistic Scholar and the Library." Library Quarterly 50 (1) January 1980: 22-39.
20. Ferguson, Chris D. "The Librarian-Historian Interface in Reference Service." RQ 21 (4) Summer 1982: 356-359.
21. Jones, C., Chapman, M., and Carr Woods, P. "The Characteristics of the Literature Used by Historians." Journal of Librarianship 4 (3) 1972: 137-156.
22. Stieg op. cit.
23. Williams, Martha E. "Feasibility Study for an Automatic Database Selector." Newsidic 20 Summer 1976: 14.
24. Tenopir, Carol. "Evaluation of Database Coverage: A Comparison of Two Methodologies." Online Review 6 (5) October 1982: 423-441.
25. Wiberly, Steven E. "Subject Access in the Humanities and the Precision of the Humanist's Vocabulary." Library Quarterly 53 (4) October 1983: 420-423.
26. Saracevic, Tefko and Kantor, Paul. "A study of information Seeking and Retrieving. II. Users, Questions, and Effectiveness." Journal of the American Society for Information Science 39 (3) May 1988: 177-196.
27. Petersen, Toni. "The Art and Architecture Thesaurus as a Front End for Subject Retrieval in Multiple Art Databases." Paper presented at the ASIS Annual Conference in Atlanta, GA., October 1988.
28. McCain, Katherine W. "Descriptor and Citation Retrieval in the Medical Behavioral Sciences Literature: Retrieval Overlaps and Novelty Distribution." Journal of the American Society for Information Science 40 (2) March 1989: 110-114.

29. Hersey, D. F. et al. "Free Text Word Retrieval and Scientist Indexing: Performance, Profiles and Costs." Journal of Documentation 27 1971: 167-183
30. Dubois, C.P.R. "Free Text versus Controlled Vocabulary: A Reassessment." Online Review 11 (4) August 1987: 167-183.
31. Perez, E. "Text Enhancement, Controlled Vocabulary versus Free Text." Special Libraries 73 (3) Summer 1982: 183-192.
32. Markey, Karen et al. "An Analysis of Controlled Vocabulary and Free Text Search Statements in Online Searches." Online Review 4 (3) September 1980: 225-236.
33. Walker, Geraldene. A Comparative Evaluation of End-user Search Performance in an Academic Environment. Ph.D. dissertation, Syracuse University, 1988.
34. Ward, Sandra N. and Osegueda, Laura M. "Teaching University Student End-users about Online Searching." Science & Technology Libraries 5 (1) Fall 1984: 17-31.